



OPINION ARTICLE

New forms of checks and balances are needed to improve research integrity [v1; ref status: indexed, <http://f1000r.es/32k>]

Elizabeth Iorns¹, Christin Chong²

¹Science Exchange Inc., 555 Bryant Street, #939, Palo Alto, CA, 94301-1704, USA

²Department of Neurology, University of California San Francisco, 1550 4th St #546A, San Francisco, CA, 94158-2324, USA

v1 First published: 28 May 2014, 3:119 (doi: [10.12688/f1000research.3714.1](https://doi.org/10.12688/f1000research.3714.1))
 Latest published: 28 May 2014, 3:119 (doi: [10.12688/f1000research.3714.1](https://doi.org/10.12688/f1000research.3714.1))

Abstract

Recent attempts at replicating highly-cited peer-reviewed studies demonstrate that the “reproducibility crisis” is indeed upon us. However, punitive measures against individuals committing research misconduct are neither sufficient nor useful because this is a systemic issue stemming from a lack of positive incentive. As an alternative approach, here we propose a system of checks and balances for the publishing process that involves 1) technical review of methodology by publishers, and 2) incentivizing direct replication of key experimental results. Together, these actions will help restore the self-correcting nature of scientific discovery.

Open Peer Review

Referee Status:

	Invited Referees		
	1	2	3
version 1 published 28 May 2014	 report	 report	 report
1	Andrew Chalmers , University of Bath UK		
2	Ivan Oransky , New York University USA, Retraction Watch USA		
3	David Soll , University of Iowa USA		

Discuss this article

Comments (0)

Corresponding author: Elizabeth Iorns (elizabeth@scienceexchange.com)

How to cite this article: Iorns E and Chong C. **New forms of checks and balances are needed to improve research integrity [v1; ref status: indexed, <http://f1000r.es/32k>]** *F1000Research* 2014, 3:119 (doi: [10.12688/f1000research.3714.1](https://doi.org/10.12688/f1000research.3714.1))

Copyright: © 2014 Iorns E and Chong C. This is an open access article distributed under the terms of the [Creative Commons Attribution Licence](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. Data associated with the article are available under the terms of the [Creative Commons Zero "No rights reserved" data waiver](#) (CC0 1.0 Public domain dedication).

Grant information: The author(s) declared that no grants were involved in supporting this work.

Competing interests: Elizabeth Iorns is employed at and owns shares of Science Exchange Inc. Christin Chong has no conflicts of interest to disclose.

First published: 28 May 2014, 3:119 (doi: [10.12688/f1000research.3714.1](https://doi.org/10.12688/f1000research.3714.1))

First indexed: 16 Jun 2014, 3:119 (doi: [10.12688/f1000research.3714.1](https://doi.org/10.12688/f1000research.3714.1))

Introduction

The scientific method provides a systematic framework for formulating, testing and refining hypotheses. By definition, it requires findings to be reliable so that theories can be refined and scientific progress can occur. Recently, it has become clear that the scientific method as it is currently being practiced is failing in self-correction, with multiple studies indicating that more than 70% of surveyed peer-reviewed articles cannot be independently verified¹⁻⁴. Unfortunately, instead of focusing on new systems to promote high quality reproducible research, most resources and attention are focused on trying to police the scientific community by investigating allegations of research misconduct. This approach is destined to fail, because the problem is systemic and not caused by a few bad players who can be caught and punished. From 1994–2003, 259 cases of misconduct were formally investigated by the Office of Research Integrity⁵. In contrast, ~480,000 papers funded by the NIH were published⁶. It would be impractical and ineffective to investigate why 70% of published findings are irreproducible, even though ultimately the ability to repeat and build upon prior work is the key component of research integrity that we should care about. Instead, truly addressing the “reproducibility crisis” requires establishing new checks and balances for the publishing process through 1) technical review of methodology by publishers, and 2) incentivizing direct replication of key experimental results. If we, the scientific community, fail to ensure the quality of the research we produce, other parties with their own vested interests will step in to police us instead⁷.

1. Checks: Publishers need to verify quality of research through third-party technical review

Publishers are uniquely placed to significantly improve reproducibility because of their inherent need to garner respect from the scientific community. *Nature* and *EMBO* are two stand-out examples who are leading the way on ensuring the quality of the research published in their journals. Moreover, current efforts to ensure quality using peer-review alone to weed out irreproducible research are not effective. One reason is that the breadth of technical knowledge that is now required to review a single study is beyond individual scientists. The number of authors per article has increased over the last decade⁸. In contrast, peer review still relies on two or three peers who are unlikely to be qualified to assess every experimental technique in the study. *Nature* has implemented an impressive new policy to reduce irreproducibility of its published papers⁹, and a key aspect to this is employing expert statisticians to review the statistical analysis of papers. Currently, a major limiting factor for implementing technical review is the lack of standardization for methodology design and required controls. Establishing and implementing these standards to ensure the technical quality of the research published in their journals is an effective value-added service that publishers should provide as a separate power in the scientific community. The Resource Identification Initiative (<https://www.force11.org/node/4463> date accessed: 2014-04-24) is an example of practical implementation for reporting of materials and methods in a standardized and machine-readable manner. Similar to successful mandates on open access to raw data, journals wield the power to require clear methodology as prerequisite for publication. Further, analogous to open data, the nascent implementation of standardized methodologies will likely yield debates, but lively discussions by the scientific community are useful for policy refinement

(<http://blogs.plos.org/everyone/2014/03/08/plos-new-data-policy-public-access-data/> date accessed: 2014-04-25).

2. Balances: Direct replication needs to be incentivized for science to be self-correcting

While journals should carry technical review responsibilities, establishing positive incentive structures for reproducible science is necessary to balance the pressure of producing high-profile publications at all costs. Of course, there will always be edge cases where it is not practical to directly replicate findings (for example unpredictable or one-off events like an earthquake), but for the majority of findings it should be possible to *directly* replicate them. That is, repeat the experiment as-is, while collecting additional information such as “*the reliability of the original results across samples, settings, measures, occasions, or instrumentation*”¹⁰. This is separate from *conceptual* replication, which is “an attempt to validate the *interpretation* of the original observation by manipulating or measuring the same conceptual variables using different techniques”¹⁰. It is also separate from re-analysis of existing raw data to check for errors in analysis and presentation, but where no new data are obtained. Therefore, directly reproducing experiments is not merely redundant effort, because new data are generated and analyzed to demonstrate the robustness of the original results.

Journals such as *F1000Research* and PLOS ONE (<http://f1000research.com/author-guidelines>, <http://www.plosone.org/static/publication>, date accessed: 2014-03-14) now consider direct replication of original studies, but even a place to publish is not sufficient because there needs to be an effective system to incentivize scientists to conduct replication studies in the first place. The simplest way to conduct replication studies is via fee-for-service technical providers because of their pre-existing methodological expertise and neutral academic involvement (i.e. they are motivated by an operational or a monetary incentive, and thus do not fear retribution from their peers or have the need to accumulate high impact ‘novel’ publications). Similarly, grants specifically designated for research integrity are vital for driving replication (<http://www.arnoldfoundation.org/reproducibility-initiative-receives-13m-grant-validate-50-landmark-cancer-studies> date accessed: 2014-04-28). These are strategies used by the Reproducibility Initiative (<https://www.scienceexchange.com/reproducibility>, date accessed: 2014-03-14), and it remains to be proven whether it will be a cost-effective mechanism to conduct direct replications.

The recent ascent of crowd-sourced post publication peer reviews have identified manuscripts with problematic content, but they remain most active for articles on new techniques that other researchers are eager to replicate for their own experiments (e.g. <http://www.ipscell.com/stap-new-data/> date accessed: 2014-04-28 and <http://f1000research.com/articles/3-102/v1> date accessed: 2014-05-20). Therefore, positively incentivizing direct replication is necessary for science to become self-correcting again, because no one would selectively publish only their experiments that worked or manipulate their findings knowing that a replication attempt, whether experimental or analytical, would not find the same significant outcome. Scientists would also be more willing to share their raw data and full methodologies before publishing because they want to make sure that their findings are reproducible. Not identifying robust and reproducible research is very costly and impairs

our ability to make effective progress against diseases like cancer in which we have already invested billions of dollars. Establishing new checks and balances with existing members of the scientific community such as publishers and fellow scientists is infinitely more preferable than those imposed by outside authorities. And if science progresses by “standing on the shoulders of giants”, it is our duty as scientists to ensure that the “shoulders” are steadfast for our peers.

Author contributions

E.I and C.C. co-wrote this article.

Competing interests

Elizabeth Iorns is employed at and owns shares of Science Exchange Inc.

Christin Chong has no conflicts of interest to disclose.

Grant information

The author(s) declared that no grants were involved in supporting this work.

Acknowledgments

We would like to acknowledge the Reproducibility Initiative board of advisors for their support.

References

1. Prinz F, Schlange T, Asadullah K: **Believe it or not: how much can we rely on published data on potential drug targets?** *Nat Rev Drug Discov.* 2011; **10**(9): 712. [PubMed Abstract](#) | [Publisher Full Text](#)
2. Begley CG, Ellis LM: **Drug development: Raise standards for preclinical cancer research.** *Nature.* 2012; **483**(7391): 531–533. [PubMed Abstract](#) | [Publisher Full Text](#)
3. Scott S, Kranz JE, Cole J, *et al.*: **Design, power, and interpretation of studies in the standard murine model of ALS.** *Amyotroph Lateral Scler.* 2008; **9**(1): 4–15. [PubMed Abstract](#) | [Publisher Full Text](#)
4. Patsopoulos NA, Tatsioni A, Ioannidis JP: **Claims of sex differences: an empirical assessment in genetic associations.** *JAMA.* 2007; **298**(8): 880–93. [PubMed Abstract](#) | [Publisher Full Text](#)
5. **ORI Closed Investigations into Misconduct Allegations Involving Research Supported by the Public Health Service: 1994–2003.** Lawrence J. Rhoades, Ph.D. [Reference Source](#)
6. Boyack KW, Jordan P: **Metrics associated with NIH funding: a high-level view.** *J Am Med Inform Assoc.* 2011; **18**(4): 423–431. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
7. Mervis J: **U.S. science policy. Bill would set new rules for choosing NSF grants.** *Science.* 2013; **340**(6132): 534. [PubMed Abstract](#) | [Publisher Full Text](#)
8. Papatheodorou SI, Trikalinos TA, Ioannidis JP: **Inflated numbers of authors over time have not been just due to increasing research complexity.** *J Clin Epidemiol.* 2008; **61**(6): 546–51. [PubMed Abstract](#) | [Publisher Full Text](#)
9. **Announcement: Reducing our irreproducibility.** *Nature.* 2013; **496**(7446): 398. [Publisher Full Text](#)
10. Collaboration, Open Science, **The Reproducibility Project: A Model of Large-Scale Collaboration for Empirical Research on Reproducibility.** 2013. [Publisher Full Text](#)

Open Peer Review

Current Referee Status:



Version 1

Referee Report 18 June 2014

doi:10.5256/f1000research.3980.r4918



David Soll

Department of Biology, University of Iowa, Iowa City, IA, USA

Iorns and Chong state in the first paragraph of their Opinion Article that “70% of surveyed peer-reviewed articles cannot be independently verified”. Iorns, who heads the company, Science Exchange, Inc., reported the same statistic in an interview with Jennifer Welsh in Business Insider, 2012. Now she and Christin Chong present a set of recommendations for alleviating this problem. But the way they support their claim that 70% of research is irreproducible is problematic. They base their value primarily on four references^{1, 2, 3, 4} that demand scrutiny. These references include three cases on drug effects that are marginal and a fourth on sex differences. Two of the references include data not peer reviewed and authored by individuals from commercial companies^{1, 2}. A third is retrospective and involves the re-evaluation of statistical calculations of the original authors³. Only one, testing the effects of drugs on increased longevity of SOD1G934 mice, provides data that can be assessed⁴, and even those data, obtained in an impressive manner, are presented in a review article.

There is merit in questioning the reproducibility of studies on marginal drug effects or sex differences, but it seems irresponsible to present, as Iorns and Chong have, a sweeping statement that 70% of all published peer-reviewed articles are irreproducible, even with the qualification of “surveyed” articles. Do these authors really believe that this 70% value applies to studies on signal transduction pathways, the phenotypes of mutants from viruses to bacteria to mammals, the interactions and roles of cytoskeletal molecules, the molecular evolution of species, the functions of molecules in embryogenesis and a vast variety of other biological fields? If Iorns and Chong had limited their commentary to the efficacy of drugs in model systems with marginal effects, they could have made an important and plausible case. But even then they would have had to do a better job referencing their argument. And to bring up the fact that 259 cases of misconduct were investigated by the Public Health Service, followed by their statement “*That in contrast ~480,000 papers funded by the NIH were published.*”, appears to be an attempt to globalize the problem by insinuation rather than hard supporting data.

The suggestion by the authors that publishers should assess the methods and statistics used by third parties is already in place. It is, obviously, the peer review system, and of course it has its problems. But the insinuation is that this process is failing in 70% of cases. Publishers should indeed be more responsible for making sure that reviewers are selected who can really assess whether the methods employed and the statistics applied are valid, especially when marginal effects are claimed. I am sure that all other scientists would whole heartedly agree with that general suggestion. But a vehicle for immediately replicating data in every published paper is extraordinarily impractical, potentially very expensive and not at all necessary in areas of research in which answers are far-more straight forward.

And who would foot the bill? The publishers? They are, in almost all cases, for-profit. For replication, they would charge a small fortune. And would scientists spend half of their research funds replicating other scientist's discoveries. With the radical decrease in funding we are now experiencing, I would not bet on it. Iorns is co-founder of Science Exchange, Inc., a for profit company that charges scientists to have measurements performed in 900 laboratories worldwide that appear to have been recruited to perform experiments for a fee, and a profit, presumably for them and a presumable cut for Science Exchange, Inc. Would Science Exchange, Inc. be the vehicle for such testing?

The authors should realize that big discoveries are immediately reproduced by other scientists, to build on those discoveries. Therefore, most scientists are obsessed with the validity of their results. And reproducibility is a tough chore if scientists do not apply the exact same procedures, under the exact same conditions, with the exact same strains and the exact same reagents. Biological systems, from cell cultures to biofilms to biochemical reactions have inherent plasticity and variability, highly responsive to the smallest changes in genetic background, temperature, composition of the atmosphere, trace elements, source of reagents and extracts, and even the quality of double distilled water. But contradictions in the results published by different laboratories have a way of "shaking themselves out". Most seasoned biologists at the bench know this is the case. Iorns and Chong have made a reasonable case for a limited area of biomedical research that involves searching for small or marginal effects and which involve apparently high noise levels. But they have presented no proof that supports their claim that 70% of all biomedical research is irreproducible, an overstatement which insinuates a significant number of scientists are at worst actively trying to dupe the rest of the scientific world or at best incompetent. By globalizing the problem to a majority of the entire scientific research community in the first paragraph of their commentary, they have sensationalized the targeted problem.

I have read this submission. I believe that I have an appropriate level of expertise to state that I do not consider it to be of an acceptable scientific standard, for reasons outlined above.

Competing Interests: No competing interests were disclosed.

Referee Report 16 June 2014

doi:10.5256/f1000research.3980.r4917



Ivan Oransky^{1,2}

¹ Department of Science, Health and Environmental Reporting, New York University, New York, NY, USA

² Retraction Watch, New York, NY, USA

Thank you for the opportunity to review this article. It makes an important argument in a critical area of inquiry, and deserves publication.

I have some specific suggestions for improvement below:

1. ***"...punitive measures against individuals committing research misconduct are neither sufficient nor useful because this is a systemic issue stemming from a lack of positive incentive."***

I'd agree that such measures are not sufficient, but what is the evidence that they are not useful?

2. ***"From 1994–2003, 259 cases of misconduct were formally investigated by the Office of Research Integrity⁵. In contrast, ~480,000 papers funded by the NIH were published⁶. It would be impractical and ineffective to investigate why 70% of published findings are irreproducible, even though ultimately the ability to repeat and build upon prior work is the key component of research integrity that we should care about."***

While it is useful to discuss ORI's limited resources, there are more recent data on their investigations, for example Figure 4 of this paper:

<http://www.plosmedicine.org/article/info%3Adoi%2F10.1371%2Fjournal.pmed.1001563>. I'd also make it clear that the ORI only has jurisdiction over fabrication, falsification, and plagiarism (FFP) – e.g. scientific misconduct -- and there is no evidence that FFP is responsible for most irreproducibility. So I wouldn't rely on ORI stats for why it's impractical and ineffective to investigate irreproducibility.

3. ***"...peer review still relies on two or three peers who are unlikely to be qualified to assess every experimental technique in the study."***

I agree, but can the authors say more about how standardization of methodology design and required controls will solve this problem?

4. ***"The recent ascent of crowd-sourced post publication peer reviews have identified manuscripts with problematic content, but they remain most active for articles on new techniques that other researchers are eager to replicate for their own experiments (e.g. <http://www.ipscell.com/stap-new-data/> date accessed: 2014-04-28 and <http://f1000research.com/articles/3-102/v1> date accessed: 2014-05-20)."***

While these two examples demonstrate cases in which post-publication peer reviews are "on new techniques that other researchers are eager to replicate for their own experiments," I'm not sure that's really where post-publication peer review is most active. I would mention PubPeer here, at the very least for context.

5. The authors make a few comments about costs, which are welcome: ***"...it remains to be proven whether it will be a cost-effective mechanism to conduct direct replications."*** and ***"Not identifying robust and reproducible research is very costly and impairs our ability to make effective progress against diseases like cancer in which we have already invested billions of dollars."*** It would be useful to try to estimate how much replication efforts will cost, and where this funding will come from.

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Competing Interests: I am the co-founder of Retraction Watch: <http://retractionwatch.com>

Referee Report 10 June 2014

doi:10.5256/f1000research.3980.r4922



Andrew Chalmers

Department of Biology and Biochemistry, University of Bath, Bath, UK

Improving reproducibility is a key challenge and topical area in the life sciences. The submitted manuscript provides a well written and interesting commentary on the topic and suggests two key approaches to improve reproducibility, based on technical review and incentivizing replication.

I think the paper is suitable for publication, but suggest the authors consider the following comments if they produce a revised version.

1. It is fair to praise Nature and EMBO's recent efforts, but many scientists would put some blame for current problems on cut down methods sections, driven by space constraints which were/are imposed by some journals such as Nature and EMBO.
2. Standardised methodologies would need to be implemented carefully to avoid stifling scientific progress in developing methods and I suggest this would need to involve scientists as well as publishers?
3. I believe that clearer and longer methods sections are an important and easily achievable way to help improve reproducibility, we have written comments on one small aspect of this, the reporting of antibody use (Helsby MA, Fenn JR and Chalmers AD (2013) Reporting research antibody use: how to increase experimental reproducibility [v2; ref status: indexed, <http://f1000r.es/1np>] F1000Research 2013, 2:153 (doi: 10.12688/f1000research.2-153.v2). The authors mention the important RII, but I suggest they could give more prominence to the importance of comprehensive reporting of methods, controls and reagents. It would link directly to their point about better technical review as this is impossible without having well documented methods.
4. The section (2) on different ways to carry out replication could more specifically mention individual scientists trying to replicate findings for their own research, this work is already carried out so involves no additional funding. The key (as mentioned) is then incentivising scientists to publish this work.
5. I wonder what the authors think of initiatives like PubMed commons, aimed at collecting comments on papers, would this provide a format for shorter comments on the ability to reproduce key findings where the scientist concerned might not feel the data warranted a full publication? Is this another useful example of crowd-sourced post publication review?

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Competing Interests: I am the founder and hold shares in CiteAb the antibody search engine.
